International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 03 March 2024
Page No: 579 - 588
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0084

Deep Learning-Based Small Face Detection from Hard Image

Sapna Shinde[1], Priti Chakurkar[2], Rashmi Rane[3]*

[1]UG - School of Computer Engineering and Technology, Dr. Vishwanath Karad MIT World Peace University, Pune, Maharashtra, India.

[2,3]Assistant Professor, School of Computer Engineering and Technology, Dr. Vishwanath Karad MIT World Peace University, Pune, Maharashtra, India.

Email id:*1032202218@mitwpu.edu.in[1], priti.chakurkar@mitwpu.edu.in[2], rashmi.rane@mitwpu.edu.in[3]*

Orcid id:*0009-0009-7284-6607*

Abstract

Facial detection usually comes first in face recognition and face analysis systems. Previously, techniques such as directed gradient histograms and cascades relied on manually-engineered features from particular photos. Nevertheless, the precision with which these techniques could identify faces in uncontrolled environments was restricted. Numerous deep learning-based face recognition frameworks have recently been developed, many of which have significantly increased accuracy, as a result of the rapid progress of deep learning in computer vision. Despite these advancements, detecting small, scaled, positioned, occluded, blurred, and faces that are partially occluded in uncontrolled conditions remains a challenge in face identification. This problem has been studied for many years but has not been completely resolved.

Keywords: *Face Detection, Region Offering Network, Deep Learning, RetinaNet.*

1. Introduction

Face detection is an important and practical issue in computer vision systems, as it serves as the initial step for various tasks such as face verification, identification, clustering, landmarks, classification, alignment, and tracking. In recent years, researchers have made significant progress in developing various techniques for face identification in real-world scenarios. However, achieving accurate and reliable face identification in challenging conditions, commonly referred to as the "wild," continues to be a complex task. This difficulty arises from several factors, including variations in facial position, occlusion, scale, lighting conditions, image quality, facial expressions, and other elements that can significantly impact the performance of face identification systems. One critical distinction between face detection and typical object detection is the scale variation involved. While object detection typically deals with relatively smaller variations in aspect ratio, face detection often encounters much larger scale alterations that can range from a few pixels to thousands of pixels. This scale variability further adds to the complexity of accurately detecting and recognizing faces in diverse settings. The conventional strategy used for early face detection efforts was to extract constructed features from the image and use several classifiers to pinpoint face regions. Two well-known research illustrate the most recent state-of-the-art advancements in face identification: support vector machines (SVM) were employed after the Haar cascade classification and the histogram of oriented gradients (HOG) were applied [13]. The WIDER FACE facial detection dataset continues to have limitations in terms of face detection accuracy in challenging photos with unresolved variances.Recent years have seen exceptional performance from deep learning—more

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 03 March 2024
Page No: 579 - 588
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0084

specifically, deep convolutional neural networks, or CNNs—in a variety of computer vision applications, such as object detection, semantic segmentation, image classification, and deep learning techniques. Some well-known benchmark evaluations, including the ImageNet Large Scale Visual Recognition Challenge (ILSVRC), are governed by these techniques. However, due to the vast number of tiny faces in the WIDER FACE dataset, dense face annotation is not feasible for these problematic face images. This highlights the distinction between humans and current face detectors and exacerbates the issue when taking the detectors' speed and memory productivity into account. Modern face detectors frequently show excellent accuracy, but because of their huge parameter sets and the methods used to manage scale robustness and contextual data, they are typically computationally and memory-intensive. However, a recent advancement in the field involves the utilization of the deep learning framework known as RetinaNet, which offers promising results in terms of recall and accuracy for facial image detection. RetinaNet addresses the limitations of traditional face detection algorithms by employing a multi-scale training process and fine-tuning the model on larger and more complex datasets, such as the WIDER FACE dataset. This approach allows the model to learn and adapt to a wide range of facial scales and variations, leading to improved performance in accurately detecting faces in challenging scenarios.

2. Experimental Methods or Methodology

Teoh et al. [1] proposed a deep learning-based methodology for designing a facial recognition and identification system. The approach involved several key steps, starting with the detection of facial regions in challenging images. To achieve this, the authors utilized the Haar feature-based cascade classifier, which has been widely used in face detection tasks. Once the faces were successfully detected, the system proceeded to the recognition phase by training a classifier. For the classifier component, Teoh et al. employed a model based on the TensorFlow framework, a popular deep-learning library. This model was trained to recognize and identify faces based on the extracted features from the detected regions.In order to deeply encode facial areas, Zhao et al.'s study [2] developed a deep neural network (DNN). They accurately located key locations on the faces by using a face alignment technique. Using Principal Component Analysis (PCA), the dimensionality of the deep features was reduced. Furthermore, the similarity of feature vectors was assessed using a combined Bayesian model, which produced extremely competitive face classification accuracy. Alghaili et al. [3] proposed a system that directly detects individuals based on various criteria by extracting and utilizing the most significant features for person recognition. They trained a Deep Convolutional Neural Network (DCNN) to extract these significant features. Subsequently, a filtering technique was employed to select the most relevant features. The minimum number that represents the identity was found by deducting the chosen features of each identity in the dataset from the features of the real image. The results demonstrated that the proposed model effectively recognized faces in different poses. Tabassum et al. [4] combined the effectiveness of the Discrete Wavelet Transform (DWT) with four different algorithms to improve face recognition (FR) accuracy. These algorithms are: (i) Principal Component Analysis (PCA) error vector; (ii) Eigenvector of Linear Discriminant Analysis (LDA); (iv) Convolutional Neural Network (CNN). The authors further integrated the four results using entropy of detection probability and a fuzzy system. The recognition accuracy was evaluated based on the images and the diversity of the database. However, the utilization of CNN led to an overfitting problem. A multi-task learning strategy is used in the research [5] to develop a single-stage neural network for face detection and recognition. It can reduce computing time by detecting and recognising several faces in a single photo. The single-stage approach has the advantages of decreased memory utilisation, a simpler inference procedure, and faster computing speed in feature extraction. [6] The six primary emotions that are

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 03 March 2024
Page No: 579 - 588
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0084

represented in human facial expressions are happiness, sadness, anger, fear, contempt, and surprise. Pre-processing face photos, extracting expression traits, and recognising facial expressions are the three primary components of facial expression recognition (FER). Pre-processing is the process of identifying and normalising faces, which is required to account for changes in lighting and position. The framework for small face detection revolves around using a VGG16 network as the backbone Convolutional Neural Network (CNN) [7]. 'conv4_3' and 'conv5_3' layer features must be combined in order to provide a detection feature map that contains both high-level and low-level semantic information. After this feature combination, dimension reduction takes place using 1×1 convolution layers, followed by a 3×3 convolution layer on the concatenated features. The output of this 3×3 convolution layer forms the final detection feature map, used for both classification and bounding-box regression. [10] It discusses the importance of face detection, face anti-spoofing, and face alignment as key components in preparing face images for FR procedures. Additionally, it mentions the use of deep learning models and various methods for improving FR accuracy. The FR system successfully handled various facial recognition (FR) attacks in different contexts. However, compared to traditional machine learning algorithms, the neural network requires a larger amount of data for training. A range of face detection and recognition techniques, from outdated approaches to modern tactics, are covered in this section. Over time, numerous techniques for face detection and recognition have been developed. These techniques can be combined with four primary object detection algorithms. Early research on face detection mostly focused on computer vision systems. One of the earliest frameworks to do real-time face detection was the Viola-Jones technique; nevertheless, in spite of further developments, there are still few useful results using this approach. Rectangular Haar-like features with a cascaded AdaBoost classifier were used in the Viola-Jones framework. However, because feature learning and classifier training are carried out independently, these techniques are not end-to-end trained. They are not as accurate as they could be, despite their speed. Support Vector Machines (SVMs), such as Haar wavelets, have been trained for face detection. Haar wavelets can differentiate between positive and negative examples for feature extraction. However, they struggle to detect faces in various poses, resulting in poor classifier performance and uncertain results. Improved localization and fewer candidates for later stages have been observed when Convolutional Neural Networks (CNNs) are used for calibration following each detection step in a cascade [8]. Each calibration stage's output is utilized to normalize the detection window's status and act as an input for the stage that follows. Facial detection algorithms that are CNN-based provide improved ways over the ones that already exist. They can be divided into two groups: one-stage approaches like RetinaNet and two-stage methods like faster R-CNN.An accurate "proposal and refinement" method is used for localization in the two-stage approach. Conversely, the one-stage approach does not rely on training principles; instead, it carefully samples scales and facial positions to produce true and false samples. Methods like sampling and reweighting are frequently employed to rectify data imbalances. The one-stage method is more accurate in localization and less prone to false positives than the two-stage method, but it is still very productive with a high recall rate.

3. Proposed Face Detection Method

It is necessary to create a unique face dataset in order to assess the correctness of the recommended approach. Furthermore, a hardware apparatus is employed to obtain the kidnapper's picture, which will be fed into the facial recognition system. The flowcharts presented in Figure 1 provide guidance for the creation of the unique face dataset methodology. A unique face dataset is being generated in order to test an ESP32-CAM that will snap photographs of people. It is important to highlight that the faces collected during the testing process will not be included in the pre-curated face

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 03 March 2024
Page No: 579 - 588
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0084

databases that are currently accessible, such the AT&T Database, which typically consists of 400 images total that represent 40 different individuals. To guarantee that the ESP32-CAM can accurately identify and evaluate a broader variety of facial characteristics than what is available in standard databases, a custom dataset is being created to take into consideration the various types of faces that might be encountered when the device is in use. Therefore, in order to achieve better accuracy, we created a proprietary face dataset for the system instead of using the publicly accessible face dataset. Once a face dataset has been constructed, the face recognition system can be trained. The images from the manually created face dataset will be stored in two folders: one for testing and the other for training. It is impossible for the identical image to appear in both folders; that is, every image of a person in both folders must have a distinct pose, different lighting, and various facial expressions. The proposed system will be trained and tested using the current photos before testing with an image directly from the ESP32-CAM.

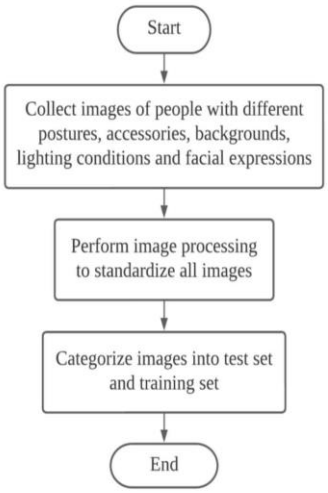

Figure 1 Flowchart of Custom Face Dataset

The ESP32CAM programming flowchart using the Arduino IDE is displayed in Fig. 2. To enable the gadget to connect to Wi-Fi or a mobile hotspot, the author [12] sets the email account that will receive the recorded photo in addition to the device's SSID and password. Deep sleep mode will be the device's first state. The smartphone will come out of deep sleep mode and start capturing images when a button is hit. The SMTP server will transmit the picture to the specified email address as soon as it is taken. The gadget will go back to deep sleep mode after the email has been successfully sent.

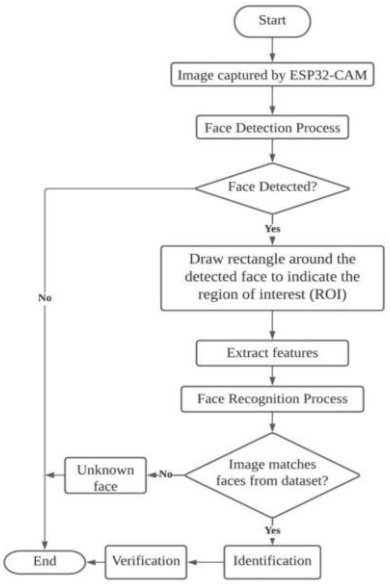

Figure 2 Face Recognition System Programming Flowchart

The work suggests a solution that makes use of the RetinaNet deep learning framework, a sophisticated design for common item detection. As seen in Fig. 3, the method consists of two fundamental components: a region offering network (RON) and a prediction branch. These components allow, respectively, the detection of faces in a given area of the picture and the development of a list of area proposals that are likely to include faces. The model can detect faces images at a competitive pace since it uses broad parameters for the convolution layers used in feature extraction. By utilising the RetinaNet architecture, the suggested method seeks to increase the recall and accuracy of facial picture detection. The training process followed the proposed systems. First, the Wider Face dataset was used to train the RetinaNet model. The same dataset was then used to assess the pre-trained model to ensure that it

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 03 March 2024
Page No: 579 - 588
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0084

produces hard negatives. By employing this trained data, fewer false positive findings can be produced in the final model. These hard negative cases were relayed over the network in the second step of the trained technique. The technique in the suggested manner was then further refined using the FDDB dataset. However, it would have been prudent to pre-train the model on a much larger face dataset with considerably more challenging cases—like the Wider Face dataset—before fine-tuning it on FDDB, given that this dataset only contains a tiny number of faces. The last level of fine-tuning involved the adoption of multi-scale training procedures.RetinaNet's end-to-end training process was chosen because to its simplicity and efficacy. The detecting bounding boxes were changed into rectangular areas of faces as a last optional step. The five essential elements that make up the suggested approach are thoroughly explained in order to enhance face image detection.

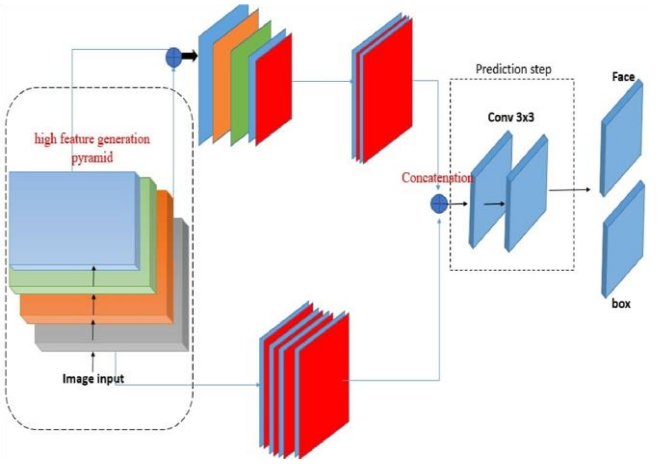

Figure 3 Diagram for A Proposed Method for Face Detection

2.1 Feature Extraction—Region Offering Network

Our network was made up of three parts. First, by merging shallow and deep features, the High Feature Generation Pyramid (HFGP) was used to produce foundational features. More specifically, multi-level semantic information for the feature maps was provided by the conv4_3 and conv5_3 layers of the ResNet architecture. Second, convolution layers were combined with a Low Feature Generation Pyramid (LFGP). These two parts were layered one after the other. The LFGP generated low-level feature maps at a different scale than the HFGP [14]. The convolution layers merged the primary features with the larger output feature map from the previous pyramid-based layers. Furthermore, the produced feature maps were sent into the subsequent convolution layer [15].

2.1.1 High Feature Generation Pyramid (HFGP)

The High Feature Generation Pyramid (HFGP) in our network combines features from one level, which is crucial for generating the final multi-scale feature pyramid. The HFGP employs 1 × 1 convolution layers on the input features' channels for compression and coupling operations to merge these feature maps. Since the HFGP receives feature maps with diverse scales from the backbone as input, it incorporates an upsampling operation to rescale the deeper features to match the scale of the coupling operation. By utilizing HFGP with very deep backbone features, stronger detection capabilities are achieved. This is because high-resolution feature maps enable better feature extraction, resulting in improved performance on small objects.

2.1.2 Low Feature Generation Pyramid (LFGP)

The LFGP is distinct from both the HFGP and RetinaNet. The composition consists of a sequence of 2-stride 3 × 3 convolution layerswhich forms a pyramid network. The input dataset for subsequent convolution layers is comprised of the feature maps from these convolution layers. In the HFGP backbone, the lower convolution layer determines the selection of the final layer at each level. After up-sampling, we implemented 1 × 1 convolution layers and described the smart sum technique within the top convolution layer network to boost learning ability and retain feature smoothness. LFGP and HFGP convolution layer outputs were combined to provide multi-scale features at the current level. The outputs of the stacked LFGP produce features that

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 03 March 2024
Page No: 579 - 588
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0084

are multi-degree and multi-scale. The initial LFGP generates shallow-level features, the second LFGP produces middle-level features, and the third LFGP generates deep-level features.

2.2 Prediction Step

The Prediction Step (PS) is formulated to integrate the multi-degree and multi-scale characteristics of the Low-Frequency Grid Patterns (LFGP) and High-Frequency Grid Patterns (HFGP) within a convolutional layer. The first step in the PS is to connect functions of the same scale collectively over the channel dimension. The resulting aggregate function pyramid can be expressed as F = [F1, F2, . . . Fi,] where Fi = Concat xi1, xi2, . . ., xiL ∈ RWi×Hi×C represents the features of the i-th large-scale. Each scale within the aggregated pyramid includes capabilities from multilevel depth. However, simple coupling operations are not sufficiently adaptive for the prediction head devoted to each feature. Therefore, we have one 3 × 3 Conv contribution via all three networks, after which every network takes its own 3 × 3 Conv in parallel. Our prediction head model is highly efficient and lightweight compared to RetinaNet.

4. Implementation and Results

Here we provide our experimental findings using a difficult dataset taken from the WIDER FACE bounding box detection challenge. We employed the WIDER FACE protocol designed to detect faces in images that pose various challenges, including occlusions, low resolution, out-of-focus faces and challenging poses. We published our findings on publicly available face datasets using the test-dev split, which is easily accessible, labelled, and does not require the usage of an assessment server, in order to compare our results with state-of-the-art methods. Lastly, for ease of comparison, we report the findings of our ablation learning trials, which were assessed on the minimal split.

4.1 Implementation Details

During our implementation, our model was trained using the PyTorch framework. For our CNN network, we selected Table 1 of ResNet 50 as the backbone, which had been pre-trained on ImageNet. Initially, we utilized the training and validation datasets from WIDER FACE as the basis for our training set. We used the levels listed in Table 1 to give each ground-truth annotation a hard value. To be more precise, we started all faces with zero problems and added a suitable hard value based on the face's location and the positive direction mentioned in Table 1. Annotations with difficulty values higher than two were not included.

To improve the accuracy of facial detection, the authors employed the technique of difficult negative mining, which involves selecting the "hard negatives" that have confidence ratings greater than 0.8 but IoU values less than 0.5 with ground-truth annotations. This technique was run for 150 iterations with a fixed learning rate of 0.0001 to ensure that these difficult negatives were included in the sample images. The FDDB dataset was used to fine tune the resulting model and was evaluated through five-fold cross-validation experiments. To prepare the face images, they were randomly resized with the shorter aspect set to 480, 600, or 750 pixels and the longer aspect capped at 1250 pixels. Throughout the training process, the researchers utilized a feature concatenation technique to combine features from the conv3_3, conv4_3, and conv5_3 layers. Furthermore, a constant scale of 4700 was consistently applied to the entire blob during both the training and testing stages. After 80 iterations with a constant learning rate of 0.001, the ultimate model was chosen. The goal of these techniques was to improve facial detection accuracy in uncontrolled conditions where small, occluded, or blurred faces can be difficult to detect.

4.2 The Process Speediness

To assess the inference speed of their model, the researchers compared it to state-of-the-art techniques. They utilized VGG-16 with its fully connected (FC) layers removed as a lighter backbone to extract base features. The inference time for each image was calculated with a batch size of 1, including the run times for both the CNN and non-maximum suppression (NMS). By adding together and dividing by 1000 the run times for 1000 face images, the average inference time per image was calculated. There were two different versions of

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 03 March 2024
Page No: 579 - 588
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0084

the model available: a conventional version consisting an input size of 512 × 512 that employed a reduced VGG16 as suggested in their study, and a fast version having an input size of 320 × 320. The multi-level structure of their approach and the benefits of one-stage detection were highlighted by their model's quick and accurate results.

Table 1 Comparison of Speed-Accuracy Curve with Other Techniques

Methods Map	Time(t)	
SSD–321	28.0	61
SSD–321	28.2	22
CornerNet	40.5	228
RetinaNet	39.1	198
RefineDet	36.7	110
FPN FRCN	36.2	172
SSD–513	31.2	125
SSD–513	31.0	37
DSSD–513	33.2	156
YOLO3–608	33.0	51
DSSA-321	28.2	85
R-FCN	29.9	85
Our	**41.0**	**84.7**

Table 1 shows the comparison of their method's speed and accuracy with other techniques. The mean average precision (mAP) and inference time are reported for each method. Their approach achieved a mAP of 41.0 and an inference time of 84.7, indicating a high level of accuracy with a competitive speed. The speed-accuracy curve demonstrated a clear positive trend compared to other approaches, further emphasizing the effectiveness of their method.

4.3 Evaluation Metrics

In our earlier research, we assessed the effectiveness of face recognition techniques using a variety of measures, including precision, recall, and FM (F-measure). The following equations illustrate that recall is the ratio of true positives to all positives in the real class, and precision is the ratio of genuine positives to all projected positives.

$$\text{Precision} = TP / (TP + FP)$$
$$\text{Recall} = TP / (TP + FN)$$

To evaluate the performance of their face recognition method, the researchers used precision, recall, and F-measure (FM) metrics. These metrics provide insights into the accuracy and completeness of the detection results.

$$FM = 2 \times \text{precision} \times \text{recall} / (\text{precision} + \text{recall})$$

The number of true positives (properly recognized faces), false positives (incorrectly detected faces), and false negatives (missed faces) are represented by the letters TP, FP, and FN in these equations. Recall is the ratio of successfully identified faces to all real faces, whereas precision is the ratio of correctly detected faces to all anticipated faces. The F-measure is a weighted average that combines precision and recall, providing a single metric to evaluate the overall performance of the face recognition method. For our proposed method, we achieved an average FM, recall, and precision of 95.6%. However, we observed a false detection rate of 4.4%, which was caused by poor lighting or images having low-quality. We also noted that wearing facial masks [11] during the COVID-19 pandemic made the process of facial recognition more challenging. Furthermore, the false positive rates of the selected methods were also evaluated in our study. Figure 4 illustrates the comparison of false positive rates, where our proposed approach demonstrated the lowest error rate among the methods examined. This indicates that our method effectively minimized the number of false positive detections. To improve the accuracy and generalization of our model, we addressed the issue of overfitting during the training process. Deep learning models frequently suffer from overfitting,

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 03 March 2024
Page No: 579 - 588
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0084

a condition in which the model becomes overly dependent on the training set and underperforms when exposed to fresh, untrained data. In order to reduce the possibility of overfitting, we used several tactics.

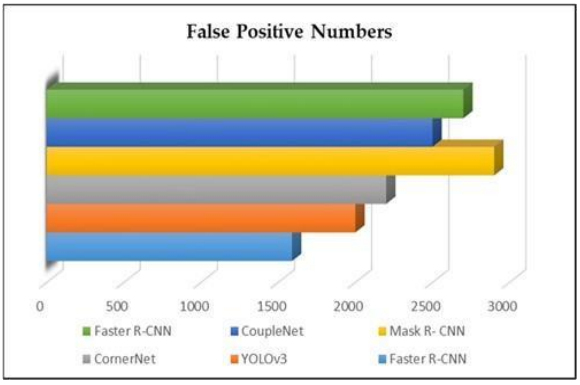

Figure 4 False Positive Feature Extraction

Initially, we employed data augmentation techniques to enhance the diversity and quantity of the training data. This involved generating additional training samples through operations such as random rotation, scaling, flipping, and noise addition. By exposing the model to a wider range of variations in the training data, we aimed to enhance its ability to generalize and perform well on unseen images. Additionally, we utilized feature selection methods to identify and retain the most relevant features for the face detection task while eliminating unnecessary or redundant features. This helped to reduce the complexity of the model and prevent it from overfitting to irrelevant or noisy features in the data. By incorporating these measures, we aimed to strike a balance between model complexity and generalization, enabling our proposed method to achieve higher accuracy and robustness in face detection while minimizing the risk of overfitting.

5. Expression Recognition Algorithm

5.1 Traditional Expression Recognition Algorithms

In classical expression recognition systems, pre-processing plays a crucial role as it helps enhance accuracy by mitigating the impact of low-quality images. Pre-processing techniques aim to eliminate irrelevant data while emphasizing relevant data. In their study, Jude, H. [9] normalized face images are then processed by the SCNN (Sketch-based Convolutional Neural Network) model, which conducts face recognition. SCNN employs a unique feature-sharing technique between shallow and hidden layers to preserve facial landmarks' localization-sensitive information. Common image pre-processing methods encompass face detection, histogram equalization, image normalization, and face alignment. To extract essential information for expression recognition from digital image signals, well-designed feature extractors are required as shown in Figure 5.

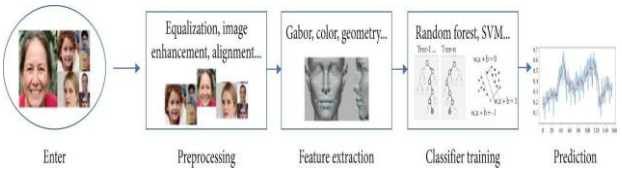

Figure 5 Traditional Facial Expression Recognition Framework

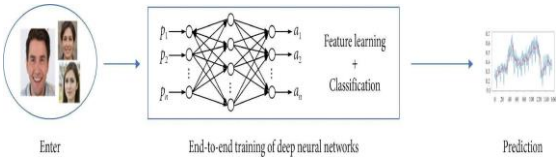

Figure 6 Deep Learning-Based Empathy Recognition Framework

5.2 Expression Recognition Algorithm Based on Deep Learning

In recent years, significant progress has been made in deep learning across various computer vision tasks, including expression recognition. Several expression recognition methods utilize deep learning techniques such as auto encoders and generative adversarial networks (GANs). Manual feature engineering is time-consuming, susceptible to inaccuracies, and constrained by specific algorithms that can only represent information within certain limitations. The domain of deep learning-driven expression recognition has evolved from small-sample tasks conducted in controlled laboratory settings to large-scale sample tasks in real-life scenarios, as depicted in Figure 6.

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 03 March 2024
Page No: 579 - 588
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0084

5.3 Expression Recognition Algorithm Based on Broad Learning

The Expression Recognition Algorithm proposed in this study leverages the innovative Broad Learning System (BLS) network, which distinguishes itself from conventional deep learning networks by adopting a distinct horizontal network structure that remains flexible throughout the training process. This unique characteristic allows the BLS network to achieve optimal classification performance. Unlike traditional deep learning networks, the BLS network exhibits a simpler structure with a reduced number of parameters, thanks to its horizontal expansion approach. The fundamental building block of the BLS network is the Random Vector Functional-Link Neural Network (RVFLNN), situated at its core. In contrast to the conventional neural network architecture, where the input data X is subjected to weight multiplication and bias addition to propagating through subsequent hidden layers, the BLS network takes a different approach based on the RVFLNN. By leveraging the advantages of the RVFLNN, the BLS network gains several benefits, including improved efficiency and simplified structure. The RVFLNN allows for efficient learning and processing of input data, contributing to the streamlined nature of the BLS network. Moreover, the horizontal expansion strategy employed by the BLS network facilitates optimal utilization of the available computational resources, enhancing its classification capabilities. Through the utilization of the BLS network and the underlying RVFLNN, the proposed Expression Recognition Algorithm achieves enhanced performance in recognizing and classifying facial expressions. By leveraging the flexibility and efficiency of the BLS network, this algorithm overcomes the limitations of traditional deep-learning approaches and offers a promising solution for expression recognition tasks.

Conclusion

In this study, we proposed a novel face-detection technique based on deep learning. Our method consists of two main components: a region-proposal network (RON) and a prediction network. The RON generates a list of potential face regions or regions of interest (RoIs), while the prediction network is responsible for classifying each RoI as a face and refining the boundaries of the detected face. By sharing the parameters of the feature extraction convolution layers, our architecture can efficiently detect faces with small model sizes and effective computation. We trained our model on the WIDER FACE dataset and evaluated its performance on both the WIDER FACE and FDDB datasets. Our proposed method has been extensively evaluated, demonstrating impressive accuracy and competitive performance compared to other one-stage detectors. The results indicate that our approach achieves an average precision (AP) of 41.0, operating at a speed of 11.8 frames per second (FPS) with a single-scale inference strategy. Moreover, by employing a multi-scale inference strategy, our method achieves an AP of 44.2, further enhancing the detection performance. The PyTorch framework was used to implement our approach, yielding a remarkable accuracy rate of 95% for successfully detected faces. However, certain challenges remain, including the detection of blurry faces in low-light conditions and the overall improvement of accuracy.To tackle these challenges, our upcoming efforts will concentrate on creating a real-time model that incorporates dependable landmark-based facial emotion recognition performance [8]. This will involve leveraging advanced techniques such as 3D Convolutional Neural Networks (3D CNN), 3D U-Net, and the YOLOv environment. Furthermore, we plan to incorporate various datasets to enhance the model's robustness and generalization capabilities. By combining these advancements, we aim to create a comprehensive solution that not only achieves real-time face detection but also delivers accurate emotion recognition. Our ultimate goal is to provide a practical and efficient system that can be deployed in various real-world applications.

References

[1]. K. H. Teoh, R. C. Ismail, S. Z. M. Naziri, R. Hussin, M. N. M. Isa, and M. S. S. M. Basir, “Face recognition and identification using deep learning approach,” in Proceedings of

International Research Journal on Advanced Engineering Hub (IRJAEH)
e ISSN: 2584-2137
Vol. 02 Issue: 03 March 2024
Page No: 579 - 588
https://irjaeh.com
https://doi.org/10.47392/IRJAEH.2024.0084

the 5th International Conference on Electronic Design, vol. 19, Perlis, Malaysia, August 2020.

[2]. F. Zhao, L. Jing, L. Z. Zhe Li, and S.-G. Na, “Multi-view face recognition using deep neural networks,” Future Generation Computer Systems, vol. 111, pp. 375–380, 2020.

[3]. Alghaili, Mohammed, Zhiyong Li, and Hamdi AR Ali. "Facefilter: face identification with deep learning and filter algorithm." Scientific Programming 2020 (2020): 1-9.

[4]. F. Tabassum, I. Islam, R. T. Khan, and M. R. Amin, “Human face recognition with a combination of DWT and machine learning,” Journal of King Saud University Computer and Information Sciences, vol. 34, no. 3, pp. 546–556, 2022.

[5]. Tsai, T.-H., Chi, Po-T,”A single-stage face detection and face recognition deep neural network based on feature pyramid and triplet loss”. IET Image Process. 16, 2148– 2156 (2022).

[6]. Mei Bie, Huan Xu, Yan Gao, Xiangjiu Che, "Facial Expression Recognition from a Single Face Image Based on Deep Learning and Broad Learning", Wireless Communications and Mobile Computing, vol. 2022, Article ID 7094539, 10 pages, 2022.

[7]. Z. Zhang, W. Shen, S. Qiao, Y. Wang, B. Wang and A. Yuille, "Robust Face Detection via Learning Small Faces on Hard Images," 2020 IEEE Winter Conference on Applications of Computer Vision (WACV), Snowmass, CO, USA, 2020, pp. 1350-1359, doi: 10.1109/WACV45572.2020.9093445.

[8]. Qingqing Xu, Zhiyu Zhu, Huilin Ge, Zheqing Zhang, Xu Zang, "Effective Face Detector Based on YOLOv5 and Superresolution Reconstruction", Computational and Mathematical Methods in Medicine, vol. 2021, Article ID 7748350, 9 pages, 2021.

[9]. Jude, H., Janarthanan, P.Jude, H.Ja ,”An Efficient Face Detection and Recognition System Using RVJA and SCNN”. Mathematical Problems in Engineering, 2022.

[10]. Wang, Xinyi, Jianteng Peng, Sufang Zhang, Bihui Chen, Yi Wang, and Yandong Guo. "A Survey of Face Recognition." arXiv preprint arXiv:2212.13038 (2022).

[11]. Hangaragia, S., ripty Singhb, Neelima Na,b, (2023). “Face Detection and Recognition Using Face Mask and Deep Neural Network”. Procedia Computer Science, Volume 218(2023), 741-749.

[12]. Jamil Abedalrahim Jamil Alsayaydeh, Irianto, Azwan Aziz, Chang Kai Xin, A. K. M. Zakir Hossain, and Safarudin Gazali Herawan, “Face Recognition System Design and Implementation using Neural Networks” International Journal of Advanced Computer Science and Applications (IJACSA), 13(6), 2022.

[13]. Anirudha B Shetty, Bhoomika, Deeksha, Jeevan Rebeiro, Ramyashree, “Facial recognition using Haar cascade and LBP classifiers”, Global Transitions Proceedings,Volume 2, Issue 2, 2021, Pages 330-335, ISSN 2666-285X, https://doi.org/10.1016/j.gltp.2021.08.044.

[14]. Mamieva, Dilnoza, Akmalbek Bobomirzaevich Abdusalomov, Mukhriddin Mukhiddinov, and Taeg Keun Whangbo. 2023. "Improved Face Detection Method via Learning Small Faces on Hard Images Based on a Deep Learning Approach" Sensors 23, no. 1: 502. https://doi.org/10.3390/s23010502

[15]. Shaoqi Hou, Dongdong Fang, Yixi Pan, Ye Li, Guangqiang Yin, "Hybrid Pyramid Convolutional Network for Multiscale Face Detection", Computational Intelligence and Neuroscience, vol. 2021, Article ID 9963322, 15 pages, 2021. https://doi.org/10.1155/2021/9963322